

Felice Dell'Orletta, Johanna Monti and Fabio Tamburini (dir.)

Proceedings of the Seventh Italian Conference on Computational Linguistics CLiC-it 2020 Bologna, Italy, March 1-3, 2021

Accademia University Press

“Spotto la quarantena”: per una analisi dell’italiano scritto degli studenti universitari via social network in tempo di COVID-19

Francesca Chiusaroli, Johanna Monti, Maria Laura Pierucci e Gennaro Nolano

DOI: 10.4000/books.aaccademia.8365

Editore: Accademia University Press

Luogo di pubblicazione: Torino

Anno di pubblicazione: 2020

Data di messa in linea: 3 settembre 2021

Collana: Collana dell'Associazione Italiana di Linguistica Computazionale

EAN digitale: 9791280136336



<http://books.openedition.org>

Notizia bibliografica digitale

CHIUSAROLI, Francesca ; et al. *“Spotto la quarantena”: per una analisi dell’italiano scritto degli studenti universitari via social network in tempo di COVID-19* In: *Proceedings of the Seventh Italian Conference on Computational Linguistics CLiC-it 2020: Bologna, Italy, March 1-3, 2021* [online]. Torino: Accademia University Press, 2020 (creato il 07 settembre 2021). Disponibile su Internet: <<http://books.openedition.org/aaccademia/8365>>. ISBN: 9791280136336. DOI: <https://doi.org/10.4000/books.aaccademia.8365>.

“Spotto la quarantena”: per una analisi dell’italiano scritto degli studenti universitari via social network in tempo di COVID-19

Francesca Chiusaroli¹, Johanna Monti², Maria Laura Pierucci¹, Gennaro Nolano²

¹Università di Macerata

²UniOr NLP Research Group, Università degli Studi di Napoli L’Orientale

f.chiusaroli@unimc.it

marialaurapierucci@gmail.com

{jmonti, gnolano}@unior.it

Abstract

Italiano Per un’analisi dell’italiano scritto degli studenti universitari, un contesto privilegiato si trova in quei canali social di condivisione tra pari noti con la titolatura di “Spotted”. Attivi particolarmente su Instagram, si tratta di canali di interazione pubblica informale che forniscono elementi significativi all’indagine sulle scritture spontanee dei giovani in rete. Questa ricerca presenta un corpus di oltre trentamila testi da Spotted universitari italiani. Concentrandoci sulle fasi COVID (e sul periodo immediatamente precedente), intendiamo illustrare come il corpus *Spotted-Ita* appaia funzionale a un’indagine riferita alla comunicazione linguistica nell’occasione di un’esperienza unica e straordinaria vissuta dagli studenti universitari.

English “Spotted” posts represent one of the most popular forms of Computer-mediated Communication (CMC) among university students in Italy, and as such, they represent a privileged context to analyze the Italian language used by students on the Web. This kind of informal communication channels is active especially on Instagram, and provides relevant insights on the spontaneous writing of young people. Based on *Spotted-ITA Corpus*, a corpus of over 30,000 posts retrieved from Italian Spotted accounts on Instagram, this paper presents a focus on the period of lockdown in Italy due to COVID pandemics and its immediately preceding pe-

riod, showing how this corpus can be useful in analyzing the way Italian university students communicated during this unique and extraordinary experience.

1 Introduzione

Tra le manifestazioni della lingua giovanile, un ruolo peculiare è assolto dalle scritture degli studenti universitari, i quali, pur nelle differenze tra i percorsi formativi (convenzionalmente distinti in umanistici e scientifici), attestano competenze di scrittura riguardanti lo strato più colto delle giovani generazioni, il più edotto sugli usi standard della lingua, e parimenti esposto alle condizioni comunicative informali del web. Su queste premesse, numerose sono le analisi e le ricerche accademiche rivolte a definire gli equilibri, le differenze, le interferenze, che si innescano tra le varie espressioni della lingua scritta all’interno del diasistema delle scritture, dai luoghi della più rigida formalità (come le tesi di laurea) a quelli della massima spontaneità (comunicazioni attraverso social network e piattaforme di messaggistica istantanea). La naturale permeabilità e l’osmotica relazione tra i livelli formale e informale, o pubblico e privato, fanno rilevare, nella scrittura dello studente, la tipica apertura a forme neologiche e non standard (per grafia, grammatica e lessico), indotte dalla cosiddetta lingua di Internet (“scritture brevi” digitali), ed evidenziano l’annessa connotazione emozionale che di tali adozioni costituisce la diretta implicazione espressiva².

²Oltre all’ampia bibliografia sul tema della lingua e dei linguaggi giovanili, di cui solo pochi titoli saranno citati nella bibliografia del presente contributo, interessa qui citare, come punto di riferimento per il tema generale, il Progetto di Ricerca Nazionale 2017, “UniverS-Ita. L’italiano scritto degli studenti universitari: quadro sociolinguistico, tendenze tipologiche, implicazioni didattiche”, attualmente in corso: (<https://site.unibo.it/univers-ita/it>). Per la nozione di “scritture brevi” digitali, cfr. Scritture Brevi, progetto di ricerca e blog, su www.scritturebrevi.it.

Per una analisi della scrittura degli studenti universitari di oggi, un contesto privilegiato si trova in quei canali social di condivisione noti con la titolarità di “Spotted”. Tali luoghi virtuali costituiscono ambiente di interazione elettivo degli studenti, in particolare per ciò che riguarda la piattaforma *Instagram*, che - rispetto ad esempio al caso di *Facebook* - garantisce all’analisi la qualità di bacheca attualmente più popolare e diffusamente utilizzata tra i giovani, idonea allo scopo di riversare i propri pensieri e anche sottoporre dubbi e quesiti indirizzandoli al gruppo di pari³.

Di generale diffusione su scala nazionale, *Spotted* costituisce un format grafico-testuale che si caratterizza innanzitutto per l’anonimizzazione dei mittenti assicurata dagli amministratori che fungono da editor al momento della pubblicazione. In questo modo, dunque, l’intervento dell’autore di ogni post appare svincolato da eventuali condizionamenti esercitati da un presidio esterno – o del mondo “adulto” (in particolare, nel caso specifico, educatori e docenti) – così che, in *Spotted*, si offre all’analisi una scrittura informale e allo stesso tempo pubblica, gergale ma intesa ad ampia diffusione, interessante per la forma oltre che per i contenuti condivisi.

Il presente contributo è così organizzato: la sezione 2 riguarda il corpus *Spotted-Ita*, sulla cui composizione si forniscono informazioni dettagliate insieme ad una descrizione delle modalità di raccolta dei post; la sezione 3 invece riguarda l’analisi linguistica condotta su due subcorpora relativi ai post pubblicati dagli studenti nei due mesi prima e nei due mesi durante il lockdown; seguono infine le conclusioni.

2 Il corpus *Spotted-Ita*

Il corpus denominato *Spotted-Ita* è costituito da 33.865 testi (“post”) pubblicati su Instagram nel

Il contributo di F. Chiusaroli e M.L. Pierucci in questo paper consiste nell’impianto teorico e di analisi e si inserisce nell’ambito del PRIN “UniverS-Ita”. J. Monti e G. Nolano hanno curato la raccolta del corpus, le collocazioni, la realizzazione dei grafici e tabelle e il più generale framework computazionale. I risultati del contributo presentato sono condivisi da tutti gli autori.

³Si vedano i report Blogmeter e Nielsen relativi all’anno in corso, disponibili online rispettivamente agli indirizzi <https://www.blogmeter.it/it/reports/italiani/-e-social-media-2020> e <https://www.nielsen.com/us/en/insights/report/2020/the-nielsen-total-audience-report-august-2020/>.

periodo che va dal 10/01/2019 al 04/05/2020⁴, per un totale di 1.063.905 token e 56.973 type.

La selezione degli account da cui raccogliere i post è stata in gran parte determinata dall’elenco degli atenei che costituiscono i soggetti della ricerca PRIN UniverS-Ita. Nello specifico, gli account *Spotted* considerati ai fini della creazione del corpus sono riportati nella Tabella 1 assieme al numero di post da essi pubblicati, mentre la Figura 1 mostra la distribuzione geografica dei post presenti nel corpus.

Account	Num.post
spotted.unibo	7.791
spotted.unicz	7.259
spotted.unimc	5.946
spotted.unina	3.024
spotted.unitorvergata	2.903
spotted.unipa	2.389
spotted.uniud	1.637
spotted.bicocca	902
spotted.unipas	670
spotted.unito	418
spotted.uniss	303
spotted.lasapienza	240
spotted.unisa	212
unimi.spotted	58
spotted.univpm	50
spotted.unica.campania	32
spotted.poliba	31
Totale	33.865

Tabella 1: Numero post per account *Spotted*.

La raccolta di dati linguistici da *Instagram* è stata particolarmente elaborata date le difficoltà intrinseche dovute alla natura dei post, che sono pubblicati sotto forma di immagini e che, per poter essere analizzati, devono essere preliminarmente convertiti in formato testuale.

Il processo di raccolta dei post è stato quindi il seguente: per scaricare le immagini e i metadati dei post interessati è stata utilizzata la libreria *Instaloader*⁵; successivamente i post sotto forma di immagini sono stati convertiti in testo attraverso il tool OCR *pytesseract*⁶. Per quanto i testi estratti con questo strumento presentino un certo margine

⁴Un ulteriore gruppo di circa 600 post, tutti pubblicati in periodo di lockdown in Italia e non inclusi nell’analisi di quest’articolo, costituisce oggetto di nuove analisi in corso sul tema. Si tratta di testi estratti da una serie di account universitari ancora di *Instagram* denominati *Poivorre* (una sorta di estensione degli *Spotted*), caratterizzati per il richiamo alla dimensione nostalgica dell’esperienza universitaria (@poivorreitorvergata usare la lavagna delle aule per studiare).

⁵<https://pypi.org/project/instaloader/>.

⁶<https://pypi.org/project/pytesseract/>.

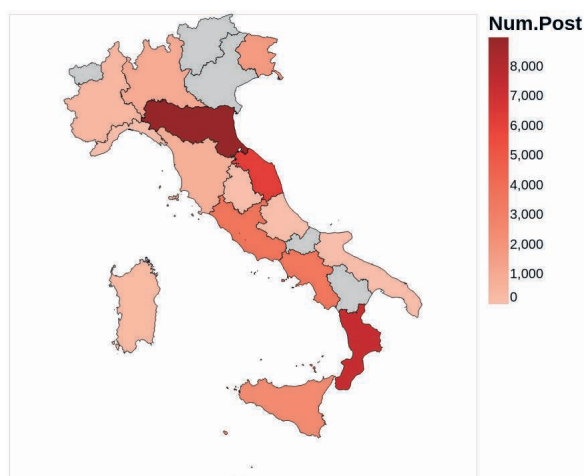


Figura 1: Mappa di calore sulla provenienza dei post *Spotted*.

di errore, sia riguardo elementi puramente testuali come punteggiatura, spazi e lettere accentate che riguardo elementi non puramente testuali (come possono essere gli emoji), tuttavia hanno mostrato un grado di leggibilità accettabile e ancorché grezzi, risultano idonei all'analisi.

È utile premettere che ciascun account *Spotted* istituisce e rappresenta comunità di scriventi idealmente delimitate dall'appartenenza ai relativi atenei del territorio nazionale.

Nonostante il carattere anonimo dei testi raccolti, la loro assimilabilità ai fini della presente analisi si istituisce dunque a partire dal criterio dell'appartenenza logicamente ipotizzabile per gli autori, ovvero l'età anagrafica, il ruolo di studenti universitari, l'appartenenza all'ateneo di riferimento dell'account, e le attività per le quali si pubblica richiesta o denuncia.

Infine un'ultima precisazione riguarda la natura dei contenuti dei post considerati per il corpus, che non sono stati filtrati per argomento. Il corpus *Spotted-Ita*, costituito in base a tali premesse, si presta ad essere interpretato a partire da molteplici prospettive di osservazione e di analisi, che molto possono dire sulla qualità dell'italiano degli studenti di livello universitario in una scrittura eterodiretta e pubblica, ma insieme libera e spontanea.

3 La lingua della pandemia: metodi per l'analisi del corpus

Una prima prospettiva di analisi del corpus è rappresentata dallo studio di come gli studenti univer-

sitari italiani abbiano vissuto e narrato il lockdown in Italia dovuto alla pandemia da COVID-19 e, di conseguenza, come questa esperienza senza precedenti si rifletta nella lingua usata nei loro post sui canali presi in esame.

A tal fine si è deciso di estrapolare due subcorpora da *Spotted-Ita*: il primo, indicato come 'pre-lockdown', copre il periodo che va dal 09/01/2020 al 07/03/2020; mentre il secondo, indicato come 'durante il lockdown', include il periodo che va dall'08/03/2020 al 04/05/2020. In tal modo, entrambi i subcorpora coprono un periodo di 58 giorni, e sono composti, rispettivamente, da 6.536 e 3.992 testi. Informazioni dettagliate sui due subcorpora (numero di token, numero di type e type/token ratio (TTR)) sono riportate nella Tabella 2.

Sezione	Num.Token	Num.Type	TTR
Pre-Lockdown	168.600	34.357	20,38%
D.Lockdown	113.185	25.449	22,48%

Tabella 2: Descrizione dei subcorpora Pre- e Durante-Lockdown.

I due subcorpora, così definiti, costituiscono l'oggetto del presente contributo, che intende soprattutto illustrare la densità del corpus, così come le potenzialità del materiale raccolto, la sua originalità sul piano dei contenuti, consentendo allo stesso tempo di verificare le consuetudini scritte favorite dal mezzo.

I due subcorpora sono stati sottoposti ad una operazione di pre-processing dei dati che ha riguardato la eliminazione di punteggiatura e stopword⁷ tramite la libreria di Python NLTK⁸, insieme a una lemmatizzazione e un tagging delle Part of Speech (PoS) con lo strumento treetaggerwrapper⁹.

Utilizzando una lista di parole chiave considerate esemplificative per il periodo preso in esame, si è proceduto all'analisi delle collocazioni¹⁰ così da analizzarne il cambiamento nei due periodi in esame. Le parole chiave possono essere suddivise nelle seguenti categorie:

⁷La lista utilizzata è quella proposta da NLTK, cui è stato aggiunto il termine 'spotto' vista l'altissima frequenza nel corpus.

⁸<https://www.nltk.org>.

⁹<https://treetaggerwrapper.readthedocs.io/en/latest/>.

¹⁰L'estrazione delle collocazioni è stata effettuata attraverso la libreria per Python NLTK, con frequenza minima = 1, e il tool online SketchEngine.

- parole della pandemia: *virus*, *Corona virus* e *corona virus* (anche unverbati), *Covid*, *lockdown*, *quarantena*;
- parole della didattica a distanza (DAD): *videolezione/i*, *lezione/i online*, *segreteria*, *Teams*, *Zoom*;
- verbi del sentimento nostalgico: *manca*, *mancano*, *mancate*, *mancare*, *tornare*.

Nella Tabella 3, vengono riportati alcuni esempi di come cambiano le collocazioni nei due periodi di riferimento presi in considerazione insieme ai relativi valori di Pointwise Mutual Information (PMI).

Termine	Pre-lockdown	D. il lockdown
covid	c. furia (11,42)	c. commovente (10,71)
	c. provvedimento (11,42)	c. god (10,71)
	epidemiologico c. (11,42)	defunto c. (10,71)
	mentana c. (11,42)	epidemiologico c. (10,71)
	c. insegnamento (10,20)	morgan c. (10,71)
quarantena	inverno q. (11,89)	angosciare q. (7,78)
	q. forzato (11,31)	combattere q. (7,78)
	q. ahahahah (11,89)	estenuante q. (7,78)
	q. volontario (10,08)	infrangere q. (7,78)
	beccare q. (9,72)	passatempo q. (7,78)
tornare	bestiame t. (10,94)	riuscire t. (9,33)
	chilometro t. (10,95)	t. delfino (9,33)
	t. canosa (10,95)	t. incazzarsi (9,33)
	t. patria (10,94)	t. patria (9,33)
	rancore t. (9,94)	t. rattusare (9,33)
mancare	m. cfu (11,79)	amiketti m. (8,23)
	finalizzare m. (10,79)	cocainomane m. (8,92)
	giustamente m. (10,79)	lisbona m. (8,92)
	m. spottedunina (10,21)	m. fiato (8,92)
	umilta m. (9,79)	m. persino (8,92)
videolezione	v. analisi (11,57)	smadonna v. (12,52)
	arrivare v. (10,43)	v. fissato (12,52)
		v. iniziare (10,93)
		v. accorgere (10,20)
		spottedlasapienza v. (9,93)

Tabella 3: Esempi di collocazioni nel corpus con indicazione del valore di PMI: la tabella è stata creata con i subcorpora normalizzati diacronicamente e lemmatizzati.

Si vedano, ne citiamo due, i seguenti post: “*Spotto il prof V. che in diretta nella pausa della lezione online ammette di aver sviluppato una dipendenza da whisky in questa quarantena*” e “*Spotto il Coronavirus e la quarantena che mi hanno fatto comprendere quanto abbia fatto bene Caino ad ammazzare quello stronzo di suo fratello!*”. È qui ben esemplificata la formula “Spotto+qualcuno/qualcosa+che” a introdurre elementi che contraddistinguono le conversazioni digitali sulla vita universitaria a distanza (*il/la prof.*, *lezione online*, *quarantena*, *Coronavirus*): si tratta di flussi grazie ai quali è possibile indagare la dimensione del *sentiment* tipicamente emergente dalla scrittura dei giovani via social, in una fase

in cui la componente emotiva risulta ancor più accentuata a causa del peculiare evento pandemico. L’incrocio dei dati fra la prospettiva temporale e quella geografica consente di osservare come, al diffondersi della pandemia e delle conseguenti restrizioni sul territorio nazionale, sia seguita una progressiva convergenza verso il tema del lockdown nell’uso della lingua in seno alle comunità afferenti alle varie sedi universitarie.

A tale riguardo, si vedano i dati riportati nelle Figure 2, 3 e 4 che mostrano le frequenze sul tempo delle parole chiave *DAD*, *nostalgia* e *pandemia* nelle tre università con frequenza maggiore di pubblicazione (rispettivamente UniBo, UniMc e UniNa), e quelli riportati nella Tabella 4, che fanno riferimento invece, alla percentuale di post *Spotted* nei quali compaiono le parole chiave *coronavirus*, *covid*, *lockdown* e *quarantena* all’interno dei due subcorpora considerati complessivamente.

Parallelamente a questo dato, che emerge nel confronto sulle tecniche pratiche (richiesta di informazioni e modalità per seguire le lezioni in teleddidattica, su tutte), si evidenzia in modo altrettanto netto l’emergenza di un sentimento “nostalgico” rilevabile dalla co-occorrenza quantitativamente marcata di lemmi riconducibili alla situazione sanitaria, come *quarantena*, con lemmi della sfera semantica delle emozioni come, ad esempio, *tornare*, *manca*, *mancano*, *mancate*, *mancare* a testimoniare il senso di solitudine e perdita (di tempo, di relazioni, di produttività negli studi).

3.1 Spotto la quarantena

Il corpus, annotato con le parti del discorso¹¹, ha inoltre permesso di effettuare ulteriori analisi relativamente all’uso ed ai contesti in cui occorre la parola *quarantena*.

Valutando il lemma nei testi estratti (cfr. Tabella 5) si nota come le co-occorrenze dalla frequenza più alta siano rappresentate dai sintagmi PRE+quarantena e PRO+quarantena, come in “in quarantena, in questa quarantena”, “durante la/questa quarantena”.

Con frequenza più bassa compaiono invece i sintagmi come “dopo questa quarantena” e “dopo la quarantena” (entrambi 5 occorrenze), e la forma semicolta “il post quarantena” (4 occorrenze), con valore temporale. Alcuni esempi dai testi:

¹¹Il tagset utilizzato è disponibile all’indirizzo <https://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/data/italian-tagset.txt>.

	gennaio	febbraio	marzo	aprile	maggio
parole della pandemia	0.22%	3.69%	10.33%	9.18%	7.44%
parole della DAD	1.95%	2.99%	18.98%	9.44%	8.68%
verbi della nostalgia	0.8%	0.74%	3.34%	5.61%	6.2%

Tabella 4: Percentuale dei post nel corpus nei quali sono presenti le parole chiave.

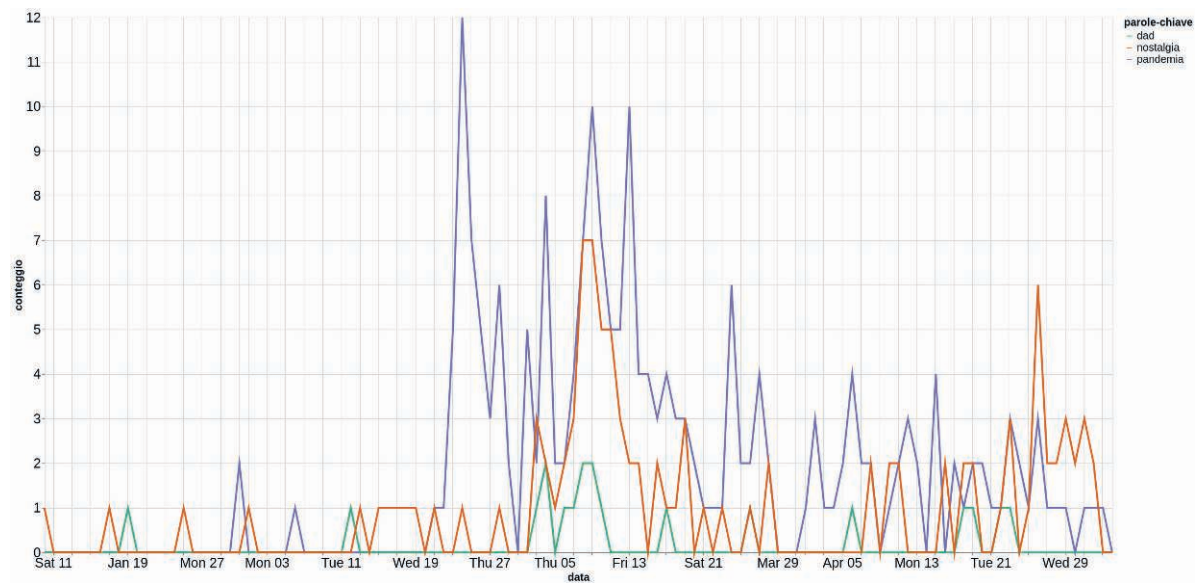


Figura 2: Frequenza delle parole chiave *DAD*, *nostalgia* e *pandemia* per i post di spotted.unibo.

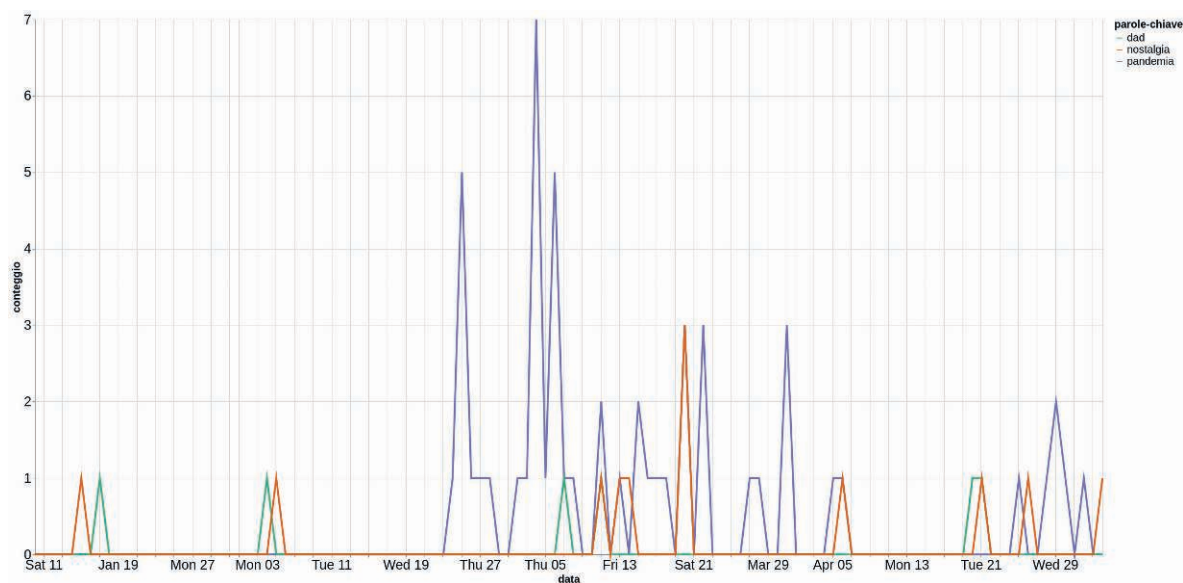


Figura 3: Frequenza delle parole chiave *DAD*, *nostalgia* e *pandemia* per i post di spotted.unimc.

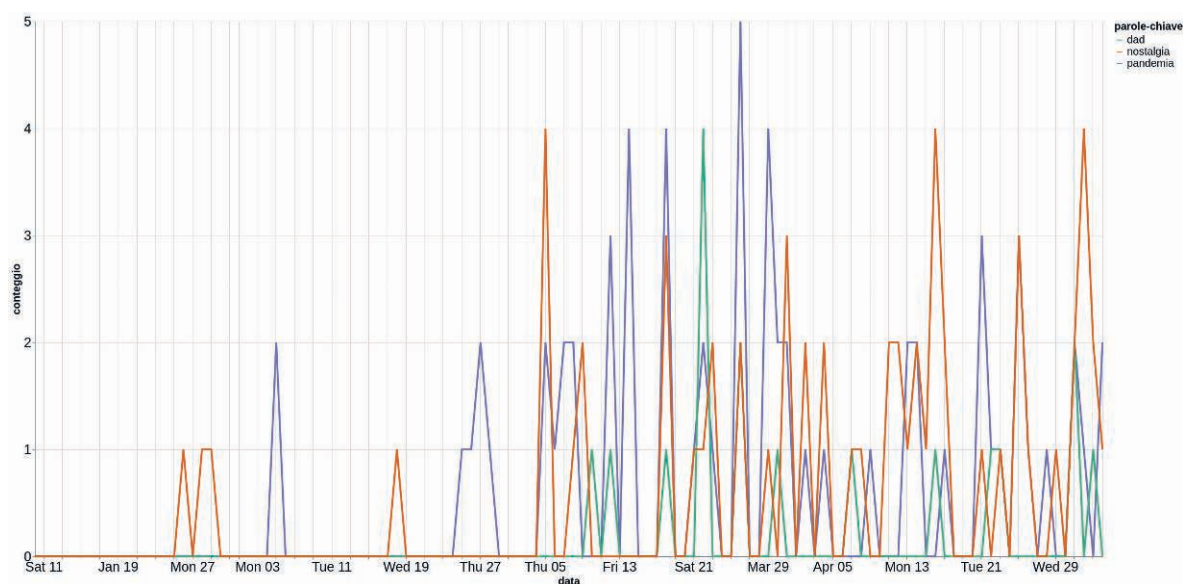


Figura 4: Frequenza delle parole chiave *DAD*, *nostalgia* e *pandemia* per i post di spottedunina.

Sintagma	Freq.
in_PRE questo_PRO:demo quarantena_NOM	34
durante_PRE il_DET:def quarantena_NOM	13
periodo_NOM di_PRE quarantena_NOM	13
giorno_NOM di_PRE quarantena_NOM	12
durante_PRE questo_PRO:demo quarantena_NOM	11

Tabella 5: PoS Tag delle co-occorrenze del termine *quarantena*.

- “spo, data la situazione di disagio complessivo che spero si capisca, cerco tipa per il **post quarantena**; garantisco ottima compagnia fornitura illimitata di risorse 420, umorismo, prestazioni sessuali e cattive intenzioni, senza contare la mia enorme capacità di guardare al futuro, dato che la tipa la cerco già da adesso... anonimo pis @ scrivi un messaggio... i) e)”
- “qualcuno disposto a due chiacchiere **in questa quarantena** forzata per tutti noi che siamo rientrati? (lo son rientrata il 24 febbraio, non faccio parte della mandria di gente tornata in queste ultime ore... incompetent!!!) comunque vi avverto che non sono una top model! dato che molti si fermano all’aspetto fisico anche per due chiacchiere”
- “organizziamo un concerto di gigione per il primo giorno **dopo la quarantena**? spotted unisa”.

I sintagmi ADJ+NOM veicolano più di altri la

connotazione di *quarantena* in termini “emotivi”: “opprimente e sterile q.”, “noioso e insopportabile periodo di q.”, “noiosa q.”.

Altri esempi:

- “spotto ragazza carina e simpatica che abbia voglia di chiacchierare per magari fare nascere qualcosa durante questa **opprimente e sterile quarantena** jç (sono un ragazzo a detta di molti simpatico e non di brutto aspetto haha) contatto i like —”
- “spotto una ragazza single con cui uscire una volta terminato questo **noioso e insopportabile periodo di quarantena!!!** contatto i likes”
- “spotto ragazze che mi tengano compagnia virtualmente in questa **noiosa quarantena**”
- “spotto qualcuno con cui fare amicizia in questa **noiosa quarantena**”.

Ancora, il termine *quarantena* funziona da catalizzatore delle conversazioni degli studenti introducendo nella scrittura stati d’animo come il senso di solitudine, la nostalgia data dalla lontananza dalle sedi universitarie e dai compagni di corso e di sede, e il desiderio del ritorno alla vita universitaria in sede. Alcuni esempi:

- “Spotto chi come me sta passando la **quarantena lontana** dalla persona che ama e che sta male per il fatto che probabilmente dopo il

4 maggio non potremo ancora spostarci tra regioni. **Non siete soli**"

- "Sotto il casino che facevamo per strada, sia a Bologna che da altre parti (NICOLOOOO), e spero che possiamo tornare a farlo ancora: Non so perche me he sono uscito. con ste cose mielose ma non mi'sto sopportando piu in questa cazzo di quarantena e **voglio tornare** al piu presto alla normalita con voi ragazzi".

Il lemma *quarantena* (328 occorrenze), termine selezionato da Treccani fra le "parole del Coronavirus"¹² nell'ambito della rubrica online "Le parole valgono", appare la verbale espressione della centralità dell'argomento COVID nel corpus di *Spotted*.

Come si vede in Figura 5, "quarantena" è entrato più tardi nell'uso rispetto allo stesso termine "coronavirus" che mostra un primo picco già alla fine di febbraio 2020. In compenso, "quarantena" risulta essere anche il termine più stabilizzato e longevo tanto da apparire con regolare frequenza fino a maggio. "Coronavirus", invece, scompare quasi del tutto dall'uso dopo i primi giorni di marzo. Interessante anche vedere come il termine più specialistico *COVID* veda un uso più raro sin dal suo primo utilizzo a febbraio, pur restando presente nella comunicazione degli *Spotted* fino alla fine del periodo preso in esame.

Nell'ottica di una valutazione della scrittura giovanile, alla quale sovente si tende a imputare la responsabilità del declino della lingua rispetto agli influssi linguistici esogeni, ci sembra qui infine interessante rilevare la pressoché nulla frequenza d'uso del termine *lockdown*, che viene impiegato in soli 2 post fra quelli raccolti nel corpus *Spotted-ita*, laddove esso risulta invece ampiamente e stabilmente adottato nello stesso periodo in ambito giornalistico e della comunicazione istituzionale.

Spostandoci sulle parole chiave della DAD, è interessante notare come il sintagma 'lezione online' veda un incremento molto basso nell'uso, come si vede dalle tabelle 6 e 7, che mostrano le frequenze dei trigrammi in cui compare il termine 'lezione'.

¹²https://www.treccani.it/magazine/parolevalgono/Le_parole_del_Coronavirus/index.html.

Sintagma	Freq.
durante_PRE il_DET:def lezione_NOM	81.0
il_DET:def lezione_NOM online_ADJ	65.0
il_DET:def lezione_NOM di_PRE	53.0
del_PRE:det lezione_NOM online_ADJ	27.0
lezione_NOM online_ADJ di_PRE	21.0
al_PRE:det lezione_NOM online_ADJ	17.0
del_PRE:det lezione_NOM di_PRE	16.0
lezione_NOM di_PRE diritto_NOM	12.0
lezione_NOM di_PRE storia_NOM	12.0
lezione_NOM online_ADJ e_CON	12.0

Tabella 6: PoS Tag delle co-occorrenze del termine "lezione" nel periodo pre-lockdown.

Sintagma	Freq.
durante_PRE il_DET:def lezione_NOM	67.0
il_DET:def lezione_NOM online_ADJ	47.0
il_DET:def lezione_NOM di_PRE	46.0
al_PRE:det lezione_NOM online_ADJ	12.0
del_PRE:det lezione_NOM online_ADJ	12.0
lezione_NOM online_ADJ di_PRE	12.0
seguire_VER:infi il_DET:def lezione_NOM	11.0
del_PRE:det lezione_NOM di_PRE	9.0
lezione_NOM online_ADJ sotto_VER:pres	9.0
sotto_VER:pres il_DET:def lezione_NOM	9.0

Tabella 7: PoS Tag delle co-occorrenze del termine "lezione" nel periodo d. il lockdown.

Nonostante si abbassino le frequenze delle forme *il_DET:def lezione_NOM online_ADJ*, *al_PRE:det lezione_NOM online_ADJ* e *lezione_NOM online_ADJ di_PRE*, compaiono trigrammi nuovi come *lezione_NOM online_ADJ sotto_VER:pres* e *del_PRE:det lezione_NOM online_ADJ*.

Alcuni esempi dai testi pre-lockdown:

- *Sotto lezione di fisica (prof M.) Ingegneria civile. Durante la lezione online e suonato l'allarme antincendio e la prof e scappata*
- *Sotto tutti quelli che si connettono 30 Min prima dell'inizio della lezione online. Avete paura di non trovare posto?.*

E alcuni esempi dai testi durante il lockdown:

- *Sotto la lezione online di economia politica da sotto le coperte con pigiamae caffè sul comodino alle 12*
- *Sotto l'eroe che durante la lezione online di diritto internazionale, accorgendosi di aver lasciato un libro nella sede di Ravenna (nella quale non puo fare ritorno) ha esclamato VACCA BOIA.*

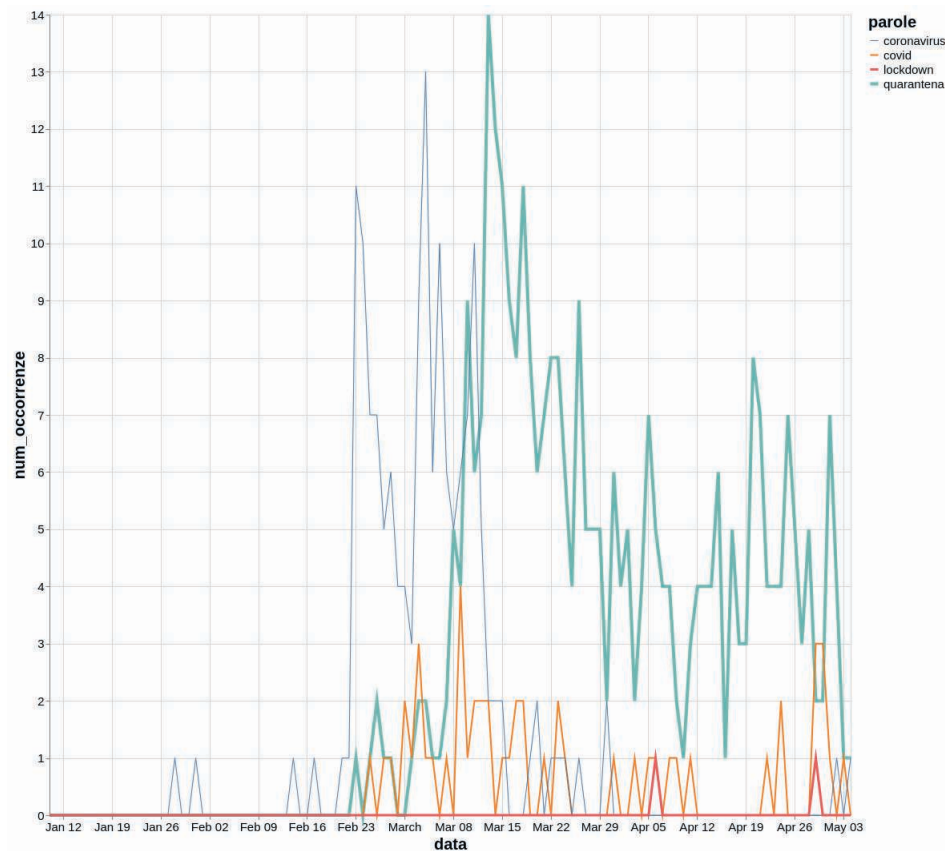


Figura 5: Conteggio delle keyword sul tempo

Infine, nell'analisi delle parole chiave della nostalgia si è notata la presenza di un fenomeno che riguarda il verbo 'mancare', che passa da 88 usi pre-lockdown a 364 durante il lockdown. Nel periodo pre-lockdown, a mancare possono essere i CFU o un/a ragazzo/a, come negli esempi seguenti:

- *Ciao spo, devo laurearmi e mi mancano Qc-fu, qualcuno mi pud dire se c'è un esame della triennale di qualsiasi facoltà molto facile? Grazie mille JI, © Scrivi un messaggio... eA*
- *Spotto Lumpa, non sei solo la migliore amica, sei la mia vita... Non posso pensare che stiamo buttando tutto via perche non abbiamo il coraggio di parlarne...mi manchi.*

Mentre, come ci si può aspettare, durante il lockdown a mancare è invece la quotidianità temporaneamente perduta:

- *Spotto la mia tremenda nostalgia di Bologna, della normalità, degli abbracci senza paura, dei baci al mio lui, della gente per la città. **Mi manca** tutto questo terribilmente, ma andrà tutto bene &*

- *Spotto chiara e melissa. Prima eravamo tutti I giorni insieme. Oggi la distanza Mantova Roma si sente. **MI mancate** tope.*
- *Spotto i miei fratelli dell'ed. 20, dal primo all'ultimo **mi mancate**. Ad alcuni quel posto mette ansia, per me era una seconda casa e spero di tornarci presto. Un bacio a tutti da 3p.*

4 Conclusioni

Analizzare l'italiano scritto degli studenti universitari attraverso i testi postati sulla piattaforma social Instagram, in particolare gli *Spotted*, fornisce un focus di ricerca nodale nel panorama degli studi accademici e scientifici sulla natura della lingua dei giovani.

In tale contesto, nel registrare l'andamento delle conversazioni nel pre- durante- e post quarantena, il corpus *Spotted-ita* fornisce un punto di vista privilegiato per l'analisi dell'esperienza dello studente nelle inedite ed esclusive condizioni conosciute nella fase sociale più significativa dell'emergenza sanitaria COVID-19.

Coadiuvata dagli strumenti della linguistica computazionale, l'analisi qui presentata illumina alcuni aspetti degli usi linguistici degli studenti, in contesti informali, tipicamente non-standard, ma altresì, per il gruppo dei pari, contesti pubblici, consentendo di ridefinire le esigenze comunicative rispetto alle nuove condizioni delle attività quotidiane e di studio svolte a distanza. Ne emerge un osservatorio peculiare e unico, dunque, che esemplarmente rispecchia e riproduce, in termini di rappresentazione linguistica, il vissuto di un periodo storico, mai altrimenti sperimentato nella contemporaneità, di disaggregazione fisica della comunità universitaria.

References

- Chiusaroli F. 2017. "Scritture brevi" nel diasistema delle scritture digitali, in C. De Santis/N. Grandi (eds.) *CLUB Working Papers in Linguistics*, Università di Bologna: 5-18.
- Cortelazzo M. 2010. Linguaggio giovanile, in *Enciclopedia dell'italiano*, disponibile all'indirizzo http://www.treccani.it/enciclopedia/linguaggio/-giovanile/_ (Enciclopedia dell'Italiano) Ultimo accesso 10/09/2020.
- D'Achille P. 2016. Architettura dell'italiano di oggi e linee di tendenza, in Lubello (ed.), *Manuale di linguistica italiana*, Berlin/Boston, De Gruyter: 165-189.
- De Mauro T. 2010. La cultura degli italiani, a cura di F.Erbani, Roma/Bari, Laterza.
- Fiorentino G. 2015. Aspetti problematici del discorso accademico: un'analisi dei riassunti delle tesi di laurea, *Cuadernos de Filología Italiana* 22: 263-84.
- Grandi N. (ed.) 2015. La grammatica e l'errore, Bologna, BUP.
- Lubello S. (ed.) 2020. Homo scribens 2.0. Scritture ibride della modernità, Firenze, Franco Cesati.
- Piemontese M.E./Sposetti P. 2014. La scrittura dalla scuola superiore all'università, Roma, Carocci.
- Palermo M. 2017 *Italiano Scritto 2.0. Testi e Ipertesti*, Roma, Carocci
- Pistolesi E. 2014. Scritture digitali, in G. Antonelli/M. Motolese/L. Tomasin (eds.), *Storia dell'italiano scritto, III: Italiano dell'uso*, Roma, Carocci: 349-75.